

深層学習を用いた地下水位モデリング

熊本大学水文研究室 学生会員 坂口大珠
 熊本大学くまもと水循環・減災研究教育センター 正会員 石田桂
 熊本大学水文研究室 学生会員 永里起義

1. はじめに

熊本県では、地下水が様々な用途で使用している。近年、水田の作付面積の減少や都市化による宅地面積の増加によって地下水涵養量が減少している。経年変化で比較すると、地下水位は減少傾向にある。また、地下水資源は世界の全飲料水のほぼ半分と灌漑用水の43%を供給している。気候変動による干ばつによって、地下水が唯一の信頼できる水資源となる地域も存在している。そのため、将来における地下水の実態を把握することが重要である。しかし、将来に関する観測データは存在しない。信頼ある将来の予測値を得るには気候変動を加味した気象データを入力し、信頼できる予測値を得る必要がある。そのため、高精度での地下水位モデリング技術の開発が求められている。

近年、深層学習と呼ばれる手法が注目を集めている。深層学習は水文分野でも適用がなされている¹⁾。深層学習は地下水位モデリングにおいて実績がある。本研究では、気象データを入力とし、高精度での地下水位モデリングを目的とする。また、推定精度が低かったものに対しては、入力変数の追加や誤差関数の変更なども検討する。

2. 手法

LSTMは深層学習手法の中でも、再帰型ニューラルネットワーク (RNN) の一種である。RNNは時系列予測や自然言語処理などに用いられている。LSTMは特徴として再帰的なメモリセルを1つ以上有する。このメモリセルはLSTMブロックと呼ばれる。LSTMブロック内には時刻 t のとき3つのゲートはそれぞれ入力ゲート $i(t)$ 、出力ゲート $o(t)$ 、忘却ゲート $f(t)$ が存在する。また時刻 t でのCell StateとHidden Stateの値はそれぞれ $c(t)$ 、 $h(t)$ である。 $g(t)$ はCellゲートである。このCellでは過去の時間のデータを記憶し、長期的な時間依存性のあるデータの学習を可能にした。LSTMの学習により調整されるパラメータは各ゲートでの重み、およびバイアスである。最後の時間ステップにおけるHidden stateを線形変換し最終的な出力を得る。LSTMブロック内の動作を表す数式は以下の通りである。

$$i(t) = \sigma(W_{ii}x(t) + b_{ii} + W_{hi}h(t-1) + b_{hi}) \quad (1)$$

$$o(t) = \sigma(W_{io}x(t) + b_{io} + W_{ho}h(t-1) + b_{ho}) \quad (2)$$

$$f(t) = \sigma(W_{if}x(t) + b_{if} + W_{hf}h(t-1) + b_{hf}) \quad (3)$$

$$g(t) = \tanh(W_{ig}x(t) + b_{ig} + W_{hg}h(t-1) + b_{hg}) \quad (4)$$

$$c(t) = f(t) \cdot c(t-1) + i(t) \cdot g(t) \quad (5)$$

$$h(t) = o(t) \cdot \tanh(c(t)) \quad (6)$$

$W_{ii}, W_{hi}, W_{io}, W_{ho}, W_{if}, W_{hf}, W_{ig}, W_{hg}$ は重み、 $b_{ii}, b_{io}, b_{if}, b_{ig}, b_{hi}, b_{ho}, b_{hf}, b_{hg}$ はバイアスである。

3. 対象地・データセット

本研究では熊本県熊本市の観測井 (約30地点) を対象地とした。地下水涵養量は約9千万立方メートルである。入力データである降水量と気温データはそれぞれAsian Precipitation - Highly Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE)から取得した。APHRODITEは観測データを元にした水平解像度が約4 kmのグリッドデータである。対象データである地下水位データは、取得先については権利の都合上非公開です。地下水位データも一時間毎の値から日平均地下水位を求めた。対象期間は2006~2015年度とした。

LSTMの実装にはPythonのフレームワークであるPyTorchを用いた。LSTMにおいて複数のハイパーパラメータを設定した。隠れ層の長さHidden Sate Length (HSL) は100を用いた。Input Data Length (IDL) は365を用いた。バッチサイズを256とし、バッチ選択にはShuffle法を用いた。誤差関数には平均二乗誤差、最適化関数にAdamを用いた。本研究ではデータセットを2006年~2011年のモデルの学習を行う訓練期間、2012年~2013年の過学習が生じていないか確認を行う検証期間、2014~2015年のモデルの最終的な評価を行うテスト期間の3つに分けた。訓練期間はモデルの学習に用いた。効率的な学習のため、Early Stoppingを用いる。Patienceは30とした。その後、テスト期間のデータを用いてモデルの最終的な精度の検証を行う。学習においては初期条件やバッチ選択にランダムな値を用いるため、ランダム性がある。そのランダム性を考慮し、観測点において、ハイパーパラメータの設定ごとに100回の学習を行った。また、地下水位の推定精度が低かったものに対し、入力変数の追加や誤差関数に変更などを検討した。入力変数についてはDay of yearを入力し、周期性を考慮した学習を行った。誤差関数に関しては自作の関数を用い外れ値を考慮しないような誤差関数を設定した。式は以下の通りである。

$$Loss = \frac{1}{n} \sum_{i=1}^n \frac{1}{4} (Q_P - Q_T) \quad (7)$$

Q_P は予測値、 Q_T は正解地である。推定流量の精

表-1 検証期間におけるNSEが最大の時の学習結果

| 観測点 | 訓練期間 | 検証期間 | テスト期間 |
|-----|--------|--------|--------|
| 上南部 | 0.802 | 0.795 | 0.752 |
| 河内 | 0.773 | 0.741 | 0.543 |
| 北部 | -1.186 | -2.824 | -0.498 |
| 池亀 | -0.15 | -5.146 | -2.533 |
| 力合 | -1.599 | -0.446 | -1.033 |
| 飽田 | 0.055 | 0.53 | -0.317 |

度評価指標として Nash-Sutcliffe Efficiency (NSE) を用いる。NSE の値は 1 から負の無限大までを示し、0.5 以上がモデルとしてよいとされている。NSE は次式で表される。

$$NSE = 1 - \frac{\sum_{t=1}^T (Q_{obs}^t - Q_{sim}^t)^2}{\sum_{t=1}^T (Q_{obs}^t - \bar{Q}_{obs})^2} \quad (8)$$

4. 結果

(1) 入力変数, 誤差関数変更なしの結果

LSTM の結果について表 - 1 に示す。なお本研究ではランダム性を考慮し、100 回の学習を行っている。議論をしやすいように検証期間で NSE が最大の時の学習結果のみを抽出し比較を行う。表 - 1 に示す通りに 30 地点のうち約 3 ケースの傾向が見受けられた。最初に、上南部、河内観測井のような地点 (Case1) では訓練期間、検証期間、テスト期間のすべての期間において NSE が 0.5 以上を示しており、高い精度で地下水位推定ができていた観測井が多くある地点である。Case1 では全期間において NSE が 0.5 以上となっていない地点も存在したが、外れ値処理を行うことで、精度が向上する可能性がある。次に、北部、池亀観測井のような地点 (Case2) では、訓練期間、検証期間、テスト期間のすべての期間において NSE が 0.5 以上を示しており、モデルとして許容される推定精度ではなかった。Case2 では、対象データにおいて急激な変化が見受けられたため、学習がうまくいっておらずこのような結果になったと考えられる。そのため、対象データに対し何らかの処理を行う必要がある。最後に、力合、飽田観測井のような地点 (Case3) では、訓練期間、検証期間、テスト期間において NSE が負の値となる地点と訓練期間、検証期間では 0.5 以上ではないが正の値を示した地点がある結果になった。Case3 では、対象データに外れ値が多いことと対象データが経年変化で増加傾向にあるため高い精度で地下水位推定ができなかったと考えられる。Case3 は入力変数、誤差関数の変更によって精度が改善する可能性がある。

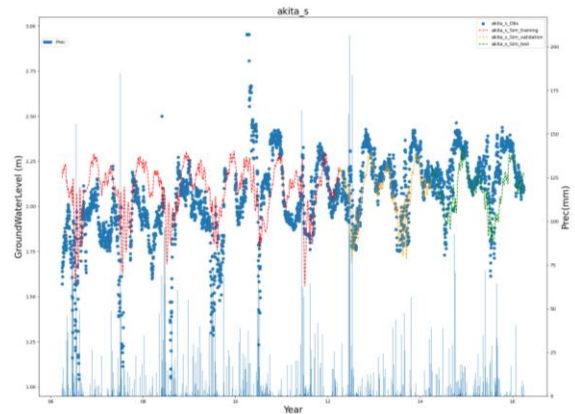


図-1 入力変数と誤差関数の変更なしのハイドログラフ (飽田地点)

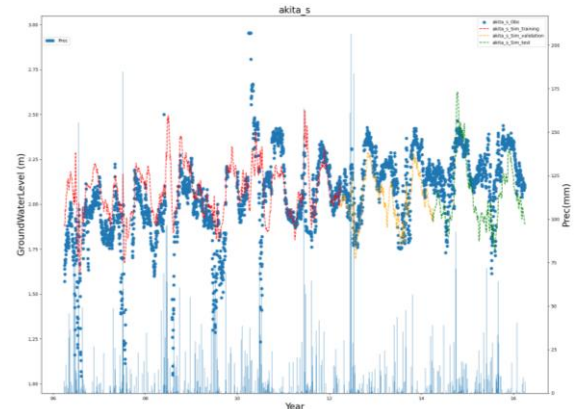


図-2 入力変数と誤差関数を変更したのハイドログラフ (飽田地点)

(2) 入力変数, 誤差関数変更した結果

Case3 である飽田観測井において入力変数と誤差関数の変更し学習した結果のハイドログラフを示す (図-1~2)。入力変数の追加、誤差関数の変更を行った結果、すべての期間の NSE は訓練期間 0.202、検証期間 0.432、テスト期間 -0.859 となった。訓練期間において NSE が向上しているが検証期間、テスト期間においては NSE が減少しており、精度改善が見受けられなかった。

5. まとめ

本研究では、LSTM を用いて気象データから地下水位推定を行った。地点によっては高精度での結果を示したが、推定精度が低い地点も存在した。入力変数、誤差関数変更したが制度の改善は見受けられなかった。今後は、外れ値の処理やハイパーパラメータの設定を変更することにより、精度を向上するための検討を行っていく必要がある。

参考文献

1) Kratzert, F. Klotz, D. Brenner, C. Schulz, K. Herrnegger. M : Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrol., Earth Syst. Sci.*, Vol.22, No.11, pp.6005-6022, 2018.